

DOCUMENT RESUME

ED 200 647

TM 810 248

AUTHOR Cohen, Peter A.
TITLE Using Student Ratings to Improve Instruction: A
Synthesis of Research Findings.
PUB DATE Apr 81
NOTE 36p.; Paper presented at the Annual Meeting of the
American Educational Research Association (65th, Los,
Angeles, CA, April 13-17, 1981).
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Feedback; Higher Education;
*Instructional Improvement; *Self Evaluation
(Individuals); *Student Attitudes; *Student
Evaluation of Teacher Performance
IDENTIFIERS Effect Size; *Meta Analysis

ABSTRACT

Meta-analytic methodology was applied to integrate findings from 24 comparisons on the effectiveness of student-rating feedback at the college level. On the average, feedback had a modest but significant effect on improving instruction. Instructors receiving mid-semester feedback averaged .14 of a rating point higher on end-of-semester overall ratings than did instructors receiving no mid-semester feedback. This corresponded to a gain of about one-third of a standard-deviation unit, or a percentile gain of 13 points. The effects of student-rating feedback were accentuated when augmentation or consultation accompanied the ratings, and were smaller when investigators controlled statistically for prefeedback ratings. Other study features, such as the length of time available to implement changes and the use of normative data, did not produce different effect sizes. (Author/RL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED200647

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Using Student Ratings to Improve Instruction:
A Synthesis of Research Findings

Peter A. Cohen

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

P. A. Cohen

Office of Instructional Services
& Educational Research
Dartmouth College
Hanover, NH 03755

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 860 248

Symposium paper presented at the Annual Meeting of the
American Educational Research Association, Los Angeles, April 1981

Abstract

This article applied meta-analytic methodology to integrate findings from 24 comparisons on the effectiveness of student-rating feedback at the college level. On the average, feedback had a modest but significant effect on improving instruction. Instructors receiving mid-semester feedback averaged .14 of a rating point higher on end-of-semester overall ratings than did instructors receiving no mid-semester feedback. This corresponded to a gain of about one-third of a standard-deviation unit, or a percentile gain of 13 points. The effects of student-rating feedback were accentuated when augmentation or consultation accompanied the ratings, and were smaller when investigators controlled statistically for prefeedback ratings. Other study features, such as the length of time available to implement changes and the use of normative data, did not produce different effect sizes.

Using Student Ratings to Improve Instruction:
A Synthesis of Research Findings

Over the last decade, there has been a dramatic increase in the use of student ratings of instruction in colleges and universities across the country. Generally, student ratings have served three functions: 1) aiding administrative evaluations of teaching effectiveness for decisions concerning pay increases, promotion, and tenure; 2) providing feedback to teachers for the purpose of improving instruction; and 3) helping students select courses and instructors. The present paper addresses the use of student ratings for improving instruction.

With the prevalent use of students as a data source for information on teaching, the body of literature concerning student ratings has become voluminous. The consensus seems to be that student ratings are reliable, valid, and not influenced to an undue extent by extraneous factors (Costin, Greenough, & Menges, 1971; Kulik & McKeachie, 1975; McKeachie, 1979; Seibert, 1979; Stumpf, Freedman, & Aguanno, 1979). While there is some reservation concerning the use of student-rating data as sufficient criteria for administrative decisions of teaching effectiveness (e.g., Cohen & McKeachie, 1980; Menges, 1979), there is little controversy over their use for purposes of improving instruction.

Efforts to improve instruction may be classified into two categories: 1) instructor development--the improvement of general teaching abilities in the instructor over time, and 2) within-class improvement--improvement in instructional effectiveness evidenced over the course of a semester.

Instructor development is evidenced by long range outcomes. For example, the instructor should show improvement in course design skills (e.g., clarity and appropriateness of course goals and objectives, measuring student attainment of objectives, etc.). Besides developing general teaching abilities, the instructor should also learn to incorporate diagnostic feedback and improvement strategies into his or her conception of the teaching-learning process. In addition, the individual's commitment to and enjoyment of teaching as a career role should be strong for those instructors who have strived to improve their teaching over a period of time.

Within-class improvement involves increased effective interaction among teacher, instructional method(s), and students throughout a semester. While the instructor can improve general teaching abilities over the course of his or her career, many teaching dimensions are relevant only in the context of a particular class of students. A large component of instructional improvement involves adapting, adjusting, and changing instruction to meet the combined needs of the students and the instructor in a particular classroom context.

In terms of providing feedback, the use of student ratings seems most appropriate for within-class improvement efforts. Cohen and Herr (1979) noted three potential advantages of using within-class improvement strategies. First, students may get increasingly better instruction as the semester progresses. At the very least, efforts to improve the course will be perceived positively by students--for example, they will feel that they have some stake in the rating process. Second, the instructor becomes actively involved in a faculty development process; not only does teaching improve but also the resistance to instructional evaluation dissipates since these strategies are seen as facilitative of improvement rather than as judgmental. Third, the intellectual and interpersonal satisfactions derived from teaching-related experiences are likely to become more salient when the instructor can overcome the frustrations associated with successfully executing the tasks of teaching. Implementation of within-class improvement strategies should increase the instructor's sense of competence.

While there seems to be substantial support for using student ratings as a source of feedback, it is not clear how effective such information can be in terms of improving instruction. A number of reviewers have noted mixed findings--some studies revealing a significant influence of student-rating feedback while other studies showing no such effects. Two recent reviews have been published dealing specifically with the effects of student evaluative feedback on instructional

improvement. However, neither review draws firm conclusions from the literature in this area. Rotem and Glasman (1979), for example, report that "feedback from student ratings (as was elicited and presented to teachers in the studies reviewed) does not seem to be effective for the purpose of improving performance of university teachers" (p.507). Using a somewhat different set of studies, Abrami, Leventhal, and Perry (1979) state, "there seems to be enough evidence to conclude that feedback from student ratings leads some instructors to improve their subsequent student ratings. However, the effect is not reliable judging from the inconsistency of findings across studies" (p. 361).

According to other reviewers in this area (Kulik & Kulik, 1974; Kulik & McKeachie, 1975; McKeachie, 1979), there seem to be four possible explanations for the failure of teachers to improve following student-rating feedback. First, the feedback should provide new information. Some studies assessing both student and self-ratings have found more improvement for instructors whose self-ratings were discrepant from their student ratings (e.g., Centra, 1973; Pambookian, 1974). Second, it may be difficult to implement changes within the short time span of an academic term. Centra's (1973) study indicates that teaching improvement may not be realized until subsequent semesters. Third, normative data may be needed to help instructors determine where their strengths and weaknesses lie. Finally, instructors may not know how

to modify their teaching once they receive student rating feedback. Consultants may play an important role here by helping to interpret feedback and suggesting change strategies.

Characteristics of the feedback implementation, which vary from study to study, may contribute to the lack of consistent findings across studies. Therefore, drawing overall conclusions from the reviews in this area proves difficult. These reviewers did not use objective, statistical methods to find the characteristics that distinguished between studies reporting substantial effects and studies reporting negligible results. In his presidential address to the American Educational Research Association, Glass (1976) proposed a method for handling the difficulties posed by the diversity of findings in the social sciences. This approach, which he called "meta analysis," involved the statistical analysis of a collection of results from individual studies for the purpose of integrating findings.

This paper applies Glass's meta-analytic methodology to research on the effectiveness of student-rating feedback for improving instruction at the college level. The results presented here update an earlier analysis by the author (Cohen, 1980). The analysis focuses on three questions. First, how effective is student-rating feedback in the typical comparative study? Second, is student-rating feedback especially effective for certain dimensions of instruction? Third, under which conditions does student-rating feedback appear to be most effective? It is intended that by addressing

these questions from a meta-analytic framework, more precise conclusions concerning the effects of student ratings on instructional improvement can be reached.

Methods

This section describes the methods that were used to locate studies, to code study features, and to quantify outcomes. The methods were similar to those developed by Kulik and his colleagues in their meta-analyses of instructional technologies in higher education (Cohen, Ebeling, and Kulik, in press; Kulik, Cohen, & Ebeling, 1980; Kulik, Kulik, & Cohen, 1979a, 1979b, 1980).

Locating Studies

The first step in the meta-analysis was to collect a number of studies that compared the effects of student-rating feedback versus no feedback on instructional improvement. Primary sources for these studies were major reviews (Abrami et al., 1979; Roten & Glasman, 1979) and four data-based indices: Psychological Abstracts, Comprehensive Dissertation Abstracts, Cumulative Index to Journals in Education and Research in Education. Secondary sources for reports on student-rating feedback were bibliographies contained in articles located through these reviews and indices.

To be included in the analysis, an article had to meet four basic criteria. First, it had to describe a study that was conducted in an actual college class. Articles on the effectiveness of student-rating feedback in secondary schools were not included (e.g., Tuckman and Oliver, 1968). Second, the study had to compare teaching improvement outcomes in

two groups of instructors--those receiving student-rating feedback and those not receiving feedback. Studies without a control group (e.g., Pambookian, 1974; Vogt & Lasher, 1973) or studies using students rather than instructors as the unit of analysis (e.g., Marsh, Fleiner, & Thomas, 1975) were not included. Third, the instructors being compared had to hold the major responsibility for teaching the class. Studies using graduate student teaching assistants (TA's) were included only if this condition was met. Fourth, the study had to be free from major methodological flaws, e.g., uncontrolled differences in initial levels of teaching ability between groups. Only comparisons that used random assignment of instructors to groups or that made statistical adjustments for differing initial teaching abilities were included. In fact, many of the studies used in the following analysis met both of these conditions.

Additional guidelines helped insure that the set of comparisons used in the analysis was as complete and representative as possible. When several papers reported the same comparison, the single, most complete report was used. When there were two distinct implementations of feedback in the same study (such as receiving consultation with the ratings or receiving ratings only), a separate comparison was obtained for each implementation. When two or more groups received the same type of feedback within a study, a single comparison was obtained.

Using these criteria, a total of 19 articles were located containing data that could be used in the meta-analysis. The 19 articles reported on 24 separate comparisons of student-rating feedback versus no feedback. The 19 studies are listed in Appendix A.

Description of Studies

The 19 studies located for this analysis took place in a variety of settings and measured outcomes in various ways. The next step in the meta-analysis was to describe the relevant study characteristics and outcomes.

Study characteristics. To characterize the studies more precisely, ten variables were defined. Two of these described methodological features--method of subject assignment and control for prefeedback ratings. Three other variables described ecological conditions under which feedback and no feedback groups were compared. These conditions included the duration of time which occurred between the feedback implementation and the outcome measure, the type of institution at which the comparison took place, and the experience of the instructors. Another three variables described the nature of the feedback. These variables specified whether or not the student-rating feedback was accompanied by consultation and/or improvement strategies, whether a self-rating discrepancy analysis was used, and whether or not normative data were included in the feedback given instructors. The final two

variables described whether or not the study was published and when. The coding categories for each of these variables and the number of comparisons in each category are listed in Table 1.

Insert Table 1 about here

Study outcomes. The next step in the meta-analysis was to express the outcomes of each comparison in quantitative terms. Outcomes used in the studies were of four major types: student ratings of the instructor, student ratings of their own learning, student attitudes toward the subject matter, and student achievement.

Student ratings of the instructor were categorized on eight dimensions. Initially, data were collected for an overall teaching effectiveness dimension. These data came from either a single rating item concerning overall teaching effectiveness or an average of all items or dimensions on a rating form in a particular study. Because instructional improvement may more likely occur for some aspects of teaching as opposed to others, rating data were also collected on seven additional dimensions of teaching. These seven dimensions were labeled Skill, Rapport, Structure, Difficulty, Interaction, Feedback to Students, and Evaluation. They are more fully described in Appendix B. All ratings for these seven dimensions and for the overall teaching effectiveness dimension were converted to a 5-point scale, where 5 represented

the highest rating (i.e., high skill, high difficulty level, etc.) and 1 represented the lowest possible rating.

The other outcomes were student ratings of their learning, student attitude toward the subject matter, and student achievement. Student ratings of their progress and their attitude toward the subject matter were converted to the same 5-point scale that was used for the instructor ratings. Student achievement was measured by performance on common final examinations.

Glass's (1976) "pure" measure of effect size was calculated as the basic index of effect for these major outcomes. Glass's ES, defined as the difference between the means for two groups divided by the standard deviation of the control group, gives the size of effect when group averages are compared. Procedures for calculating effect size were derived from McGaw and Glass (1980).

Results

This section reports the results of two different analyses. The first analysis examined the overall size and significance of effects of student-rating feedback. The second analysis was conducted to determine whether reported effects of student-rating feedback were different for different types of studies and under different conditions.

Overall Effects

In the first set of analyses, simple descriptive statistics were used to compare the results of student-rating feedback versus no feedback. Results were compiled on four outcome criteria: a) student ratings of the instructor; b) student ratings of their progress; c) student attitude toward the subject matter area; and d) student achievement. Table 2 presents the means, average effect sizes, and the results of a statistical analysis in which each comparison was treated as a single case for these outcome measures.

Insert Table 2 about here

Student ratings of the instructor. The feedback group received higher end-of-term "total" ratings than the no feedback group in 21 of the 24 comparisons. A total of 10 of the 24 comparisons reported statistically significant

differences between feedback and no feedback groups and in each case, the comparison favored the feedback group.

Continuous measures of effect size permit a more exact description of the influence of student-rating feedback on end-of-term ratings. As shown in Table 2, the average total rating was 3.89 for the feedback groups; the average total rating was 3.75 for the no feedback groups. Total end-of-term ratings for feedback and no feedback groups differed, therefore, by .14 of a rating point (on a 5-point scale); the standard deviation of this difference was .17. It is statistically very unlikely ($p < .001$) that a difference of this size would be found if there was no overall difference in effectiveness of receiving or not receiving mid-term student-rating feedback.

The median ES for the 24 comparisons was .32. Thus, the effect of student-rating feedback in a typical comparison was to raise end-of-term ratings by almost one-third of a standard-deviation unit. This implies that a typical instructor receiving student-rating feedback was performing at the 63rd percentile (as measured by end-of-term "total" rating), whereas the typical control group instructor performed at the 50th percentile. Cohen (1977) described effects of this magnitude as small to medium. This is to be contrasted with small ($ES = .2$), medium ($ES = .5$), and large ($ES = .8$) effects.

Table 2 also presents results of the analyses concerning data collected on the seven dimensions of student ratings of instruction. For four of these dimensions--Skill, Rapport, Feedback to Students, and Evaluation--mid-term student-rating

feedback had a statistically significant impact on end-of-term ratings. For the dimensions of Structure, Difficulty, and Interaction, however, mid-term feedback did not lead to statistically significant increases in end-of-term ratings.

Other outcome measures. Student Progress ratings measured students' perception of their learning or progress in a course. Four studies (five comparisons) reported data for this outcome. While only one comparison was statistically significant, all five comparisons favored the feedback group. Continuous measures of effect size showed that students whose instructors received mid-term feedback rated their own learning significantly higher than students whose instructors did not receive feedback. The median ES for this measure was .31, a small to modest effect.

Three studies (four comparisons) presented data on student attitudes toward the subject area. For all four comparisons, differences in student attitudes toward the subject favored the feedback group. In two of these comparisons, the difference was statistically significant. For these four comparisons, continuous measures of effect size showed that students whose instructors received mid-term feedback were significantly more positive in their attitudes toward the subject matter than students whose instructors did not receive feedback. This effect corresponded to an ES of .44.

Student achievement data were available from three studies (four comparisons). Not enough information was

presented in these studies to convert achievement data to a common metric (i.e., difference in examination score between the two groups, expressed as a percentage). Three of the four comparisons showed achievement to be greater for students whose instructors received feedback, while one comparison favored students whose instructors did not receive feedback. In none of the comparisons was the difference statistically significant. For this set of comparisons ES averaged .49.

Study Characteristics and Study Outcomes

The purpose of the second set of analyses was to determine whether the comparisons that reported large effects differed systematically from those reporting small effects. Table 3 presents the correlations between total rating effect sizes and study characteristics.

Insert Table 3 about here

The table shows that only one variable--augmentation of feedback--was significantly related to effect size. The median ES was .56 for comparisons that used some sort of augmentation or consultation in conjunction with student rating feedback. Effects were smaller when ratings only feedback was used. The median ES was only .23 for these comparisons. Thus, a typical instructor receiving augmented feedback was performing (at the end of the semester) at the

71st percentile. This can be compared with the typical instructor receiving only student ratings (59th percentile) and the typical instructor receiving no mid-semester feedback (50th percentile). Effect size comparisons for the other seven rating dimensions are shown in Table 4.

Insert Table 4 about here

Other study features were less highly related to effect size. None of the correlations between effect size and the remaining variables could be considered significantly different from zero. To investigate the possibility that a combination of study features might predict effect sizes more accurately than a single predictor, a stepwise multiple regression analysis was conducted. One other variable--statistical control for prefeedback differences--entered into the regression equation. Effect sizes tended to be larger (i.e., inflated) when investigators did not control statistically for mid-term differences in ratings.

Discussion

This meta-analysis showed that for the most part student-rating feedback has made a modest but significant contribution to the improvement of college teaching. These findings do not totally support the conclusions of recent reviewers in this area. For example, Abrami et al. (1979) and Rotem and Glasman (1979) have suggested that although student-rating feedback may be useful for some instructors, it has had little overall impact on increasing instructional effectiveness. The use of meta-analytic techniques, however, makes it possible to reach more exact conclusions about the effects of student-rating feedback.

The first set of analyses reported on the overall size and significance of effects of student-rating feedback. In the typical implementation, feedback raised instructors' end-of-term overall rating by .14 of a rating point, or a percentile gain of 13 points. In addition to this overall effect, feedback led to increased end-of-term ratings for four of the seven teaching dimensions. The largest effect sizes were found for the Skill, Feedback to Students, and Evaluation dimensions. As on the overall rating, the effect of mid-term student-rating feedback was to raise instructors' end-of-term ratings on these dimensions by about one-third of a standard deviation.

It is interesting to note that feedback does not lead to significant increases in ratings for all dimensions of

teaching. Ratings on some dimensions, for example Difficulty and Interaction, may be more influenced by course setting characteristics (e.g., subject matter differences, class size) and therefore not a true reflection of an instructor's effectiveness. It has also been suggested (Cohen & Herr, 1979; Sherman & Winstead, 1975) that feedback must be of a specific nature to be useful. Ratings on dimensions such as Skill or Feedback to Students provide the instructor with specific information, giving more direction to possible instructional changes.

Data for the other outcome measures were not reported as fully by the studies used in this meta-analysis. Only five comparisons were available for student progress ratings, and only four comparisons were located for attitude toward the subject and student achievement. Students whose instructors received mid-term rating feedback may have learned more than students whose instructors did not receive feedback. This is evidenced by both students' self-report of learning and their scores on achievement measures. There also is some indication that students whose instructors received feedback rated the subject-matter area higher than did students whose instructors did not receive feedback. Perhaps a noticeable improvement effort on the part of the instructor generates more student enthusiasm for the subject matter. These results, though, are at best suggestive because of the small number of available comparisons.

The second set of analyses reported on the relationship of study features and study outcomes. In general, there was little relationship between study features and study outcomes. Almost all studies randomly assigned instructors to feedback groups. Likewise, almost all feedback implementations occurred within one academic term. Findings were the same for different types of schools and for instructors with differing amounts of experience. Concerning the nature of the feedback, neither comparisons using self-rating/student rating discrepancies nor those using normative data produced differing results. Finally, whether or not the study was published and year of publication made little difference in outcomes.

The most important variable that correlated with total rating outcome was use of augmentation of feedback. Comparatively large effect sizes emerged from studies using augmented feedback or consultation in conjunction with student-rating feedback. Studies using only student-rating feedback produced much smaller effects. These results clearly suggest that instructors need more than just student-rating feedback to markedly improve their instruction.

The studies by Erickson and Erickson (1979), McKeachie et al. (1980), and Overall and Marsh (1979) are important because these investigators reported especially strong effects of consultation in conjunction with student-rating feedback. In each of these implementations, the consultants were "experts"

in the area of college teaching. As in many of the other studies using consultation, improvement goals and change strategies were discussed.

In at least one respect, the meta-analysis does not confirm the impressions of other reviewers concerning the relationship of student-rating feedback to instructional improvement. The present analysis shows that the use of normative data does not seem to enhance instructional improvement. Comparisons with one's colleagues perhaps belongs with a more evaluative use of student ratings. For formative purposes, personal norms may be more appropriate.

What implications do these findings have for using student-rating feedback to facilitate within-class improvement? It is evident that when instructors are left to their own resources, ratings provide little help. Augmented feedback, or more specifically expert consultation, seems to be the key element for making student-rating data useable for improvement purposes. It is also clear that instructional change cannot be accomplished on all teaching dimensions. Therefore, instructors should only request mid-term student feedback for aspects of teaching that they are able to modify. Finally, rating items and dimensions that supply specific information should be preferred to global ratings. With these considerations, student ratings are a valuable data source for improving instruction at the college level.

References

- Abrami, P. C., Leventhal, L. & Perry, R.P. Can feedback from student ratings help to improve college teaching? Proceedings of the 5th International Conference on Improving University Teaching. London, 1979.
- Aleamoni, L. M. The usefulness of student evaluations in improving college teaching. Instructional Science, 1978, 7, 95-105.
- Braunstein, D. N., Klein, G. A., & Pachla, M. Feedback expectancy and shifts in student ratings of college faculty. Journal of Applied Psychology, 1973, 58, 254-258.
- Carter, K. R. The effect of student feedback in modifying teaching performance. Unpublished doctoral dissertation, University of Georgia, 1974.
- Centra, J. A. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65, 395-401.
- Cohen, J. Statistical power analysis for the behavioral sciences (Rev. ed.). New York: Academic Press, 1977.
- Cohen, P. A. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 1980, 13, 321-341.
- Cohen, P. A., Ebeling, B. J., and Kulik, J. A. A meta-analysis of outcome studies of visual-based instruction. Educational Communication and Technology Journal, in press.

- Cohen, P. A., & Herr, G. Using an interactive feedback procedure to improve college teaching. Hanover, NH: Dartmouth College, Office of Instructional Service and Educational Research, February 1981.
- Cohen, P. A., & Herr, G. A procedure for diagnostic instructional feedback: The Formative Assessment of College Teaching (FACT) model. Educational Technology, 1979, 19 (12), 18-23.
- Cohen, P. A., & McKeachie, W. J. The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 1980, 28, 147-154.
- Costin, F., Greenough, W. T., & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.
- Erickson, G. R., & Erickson, B. L. Improving college teaching: An evaluation of a teaching consultation procedure. Journal of Higher Education, 1979, 50, 670-683.
- Froman, R. D., & Owen, S. V. Influence of different types of student ratings feedback upon later instructional behavior. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.
- Glass, G. V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.
- Howard, G. S. A program to improve instruction: A promising area for psychologists. Professional Psychology, 1977, 8, 316-327.
- Hoyt, D. P., & Howard, G. S. The evaluation of faculty development programs. Research in Higher Education, 1978, 8, 25-38.

- Kulik, J. A., Cohen, P. A., & Meeling, B. J. Effectiveness of programmed instruction in higher education: A meta-analysis of findings. Educational Evaluation and Policy Analysis, 1980, 2 (6), 51-64.
- Kulik, J. A., & Kulik, C-L. C. Student ratings of instruction. Teaching of Psychology, 1974, 1, 51-57.
- Kulik, J. A., Kulik C-L. C., & Cohen, P. A. A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 1979, 34, 307-318. (a)
- Kulik, J. A., Kulik, C-L. C., & Cohen, P. A. Research on audio-tutorial instruction: A meta-analysis of comparative studies. Research in Higher Education, 1979, 11, 321-341. (b)
- Kulik, J. A., Kulik, C-L. C., & Cohen, P. A. Effectiveness of computer-based college teaching: A meta-analysis of findings. Review of Educational Research, 1980, 50, 525-544.
- Kulik, J. A., & McKeachie, W. J. The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), Review of research in education (Vol. 3). Itaska, IL: F. E. Peacock Publishers, 1975.
- Marsh, H. W., Fleiner, H., & Thomas, C. S. Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology, 1975, 67, 833-839.
- McGaw, B., & Glass, G. V. Choice of the metric for effect size in meta-analysis. American Educational Research Journal, 1980, 17, 325-337.

- McKeachie, W. J. Student ratings of faculty: A reprise. Academe, 1979, 65, 384-397.
- McKeachie, W. J., Lin, Y-G., Daugherty, M., Moffett, M. M., Neigler, C., Nork, J., Walz, M., & Baldwin, R. Using student ratings and consultation to improve instruction. British Journal of Educational Psychology, 1980, 50, 168-174.
- Menges, R. J. Evaluation teaching effectiveness: What is the proper role for students? Liberal Education, 1979, 65, 356-370.
- Miller, M. T. Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 1971, 62, 235-239.
- Murphy, J. B. & Appel, V. H. The effect of mid-semester student feedback on instructional change and improvement. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, March 1978.
- Overall, J. U. & Marsh, H. W. Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 1979, 71, 856-865.
- Pambookian, H. S. The initial level of student evaluation on instruction as a source of influence on instructor change after feedback. Journal of Educational Psychology, 1974, 66, 52-56.
- Payne, D. A., & Hobbs, A. M. The effect of college course evaluation feedback on instructor and student perceptions of instructional climate and effectiveness. Higher Education, 1979, 8, 525-533.

- Rotem, A. The effects of feedback from students to university professors: An experimental study. Unpublished doctoral dissertation, University of California, Santa Barbara, 1975.
- Rotem, A. & Glasman, N. S. On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research, 1979, 49, 497-511.
- Scott, N. H. A pilot study testing effects of feedback on teaching behavior as measured by student ratings at New River Community College, Dublin, Virginia. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, 1976.
- Seibert, W. F. Student evaluations of instruction. In S. C. Erickson (Ed.), Support for teaching at major universities. Ann Arbor, MI: The University of Michigan, Center for Research on Learning and Teaching, 1979.
- Sherman, T. M. & Winstead, J. C. A formative approach to student evaluation of instruction, Educational Technology, 1975, 15, 34-39.
- Smith, D. L. The relationship of feedback to professors of the results of student ratings of their teaching effectiveness at mid-semester to their end-of-semester ratings. Unpublished doctoral dissertation, University of Southern California, 1977.
- Stumpf, S. A., Freedman, R. D., Aguanno, J. C. A path analysis of factors often found to be related to student ratings of teaching effectiveness. Research in Higher Education, 1979, 11, 111-123.

- Thomas, H. B. Jr. Changing teacher behavior: An experiment in feedback from students to teachers. Unpublished doctoral dissertation, Purdue University, 1969.
- Tuckman, B. W. & Oliver, W. F. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 1968, 59, 297-301.
- Vogt, K. E. & Lasher, H. Does student evaluation stimulate improved teaching? Bowling Green, Ohio: Bowling Green State University, 1973 (ERIC Document Reproduction Service No. ED 078 748).
- Weerts, R. R. The use of feedback from student ratings for improving college teaching. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada, March 1978.

TABLE 1

CATEGORIES FOR DESCRIBING STUDIES
AND NUMBER OF STUDIES IN EACH CATEGORY

Coding Categories	Number of Comparisons
<u>Methodological Features</u>	
Random assignment of comparison groups	
1. No	2
2. Yes	22
Statistical control for prefeedback ratings	
1. No	11
2. yes	13
<u>Ecological Conditions</u>	
Duration	
1. One term	21
2. More than one term	3
University Setting	
1. Comprehensive, liberal arts or community college	5
2. Doctorate-granting institution	19
Instructor experience	
1. Teaching assistant	9
2. Faculty member	15
<u>Nature of the Feedback</u>	
Type of use	
1. Ratings only	14
2. Ratings and augmented feedback/consulting	10
Use of self-rating discrepancies	
1. No	16
2. Yes	8
Use of normative data	
1. No	13
2. Yes	11
<u>Publication Features</u>	
Source of study	
1. Unpublished	11
2. Published	13
Study Year	
1. Before 1970	1
2. 1970-1975	5
3. 1976-1981	18

TABLE 2
MEAN RATINGS AND EFFECT SIZES
FOR MAJOR OUTCOME MEASURES

Outcome	Number of Comparisons	Means		t	Median Effect Size
		Feedback	Control		
Total Rating	24	3.89	3.75	4.03***	.32
Skill	17	3.96	3.80	5.19***	.33
Rapport	13	4.06	4.00	2.22*	.23
Structure	12	3.76	3.61	1.68	.15
Difficulty	6	2.44	2.40	.95	.14
Interaction	13	4.06	3.91	2.04	.20
Feedback to Students	7	3.65	3.49	2.93*	.32
Evaluation	6	4.02	3.84	3.00*	.33
Student Progress	5	3.88	3.76	5.66**	.31
Attitude toward Subject	4	3.30	3.15	4.95*	.44
Student Achievement	4				.49

*p < .05
**p < .01
***p < .001

TABLE 3

CORRELATIONS OF STUDY CHARACTERISTICS
WITH TOTAL RATING EFFECT SIZE

Study Characteristic	Correlation with Effect Size
Random assignment of comparison groups	-.02
Statistical control for prefeedback ratings	-.19
Duration	.08
University setting	.15
Instructor experience	.01
Use of augmentation of feedback	.62*
Use of self-rating discrepancies	.31
Use of normative data	.18
Source of study	.32
Publication year	.36

* $P < .01$

TABLE 4
COMPARISON OF RATING EFFECT SIZES FOR
AUGMENTED AND RATINGS ONLY GROUPS

Rating Dimension	Average Effect Size	
	Augmented	Ratings Only
Total Rating	.56	.23
Skill	.58	.24
Rapport	.32	.03
Structure	.32	.10
Difficulty	.02	.26
Interaction	.55	-.05
Feedback	.43	.32
Evaluation	.50	.32

Appendix A
Studies Used in the Meta-Analysis

<u>Study</u>	<u>Number of Comparisons</u>
1. Aleamoni (1978)	1
2. Braunstein, et al. (1973)	1
3. Carter (1974)	1
4. Centra (1973)	2
5. Cohen and Herr (1981)	2
6. Erickson and Erickson (1979)	1
7. Froman and Owen (1980)	1
8. Howard (1977)	1
9. Hoyt and Howard (1978)	1
10. McKeachie et al. (1980)	2
11. Miller (1971)	1
12. Murphy and Appel (1978)	2
13. Overall and Marsh (1979)	1
14. Payne and Hobbs (1979)	1
15. Rotem (1975)	1
16. Scott (1976)	1
17. Smith (1977)	1
18. Thomas (1969)	1
19. Weerts (1978)	2

Appendix B

Rating Dimensions Used in the Meta-Analysis

- 1) Skill. The Skill dimension represents the overriding quality to which students respond when rating instructors. Typical items are: "The instructor has a good command of the subject matter"; "The instructor gives clear explanations"; and "The instructor teaches near the class level."
- 2) Rapport. The Rapport dimension includes items dealing with a teacher's empathy, friendliness, approachability, and accessibility. Sample items are: "The instructor is friendly"; "The instructor is permissive and flexible"; and "The instructor is available to talk with students outside of class."
- 3) Structure. The Structure dimension describes how well the instructor planned and organized the course. Typical items are: "The instructor has everything going according to schedule"; "The instructor uses class time well"; and "The instructor explains course requirements."
- 4) Difficulty. The Difficulty dimension deals with the amount and difficulty of the work the teacher expects of students. Typical items are: "The instructor assigned difficult reading"; "The instructor asked for more than students could get done"; and "This course required more work than others of comparable credit hours."
- 5) Interaction. The Interaction dimension measures the degree to which students are encouraged to share their ideas and become actively involved in class sessions. Typical Interaction items are: "The instructor encourages students to express various points of view"; "The instructor encourages students to volunteer their own opinions"; and "The instructor facilitates classroom discussion."
- 6) Feedback. The Feedback dimension measures the instructor's concern with the quality of students' work. Standard items for this dimension are: "The instructor tells students when they have done a particularly good job"; "The instructor checks to see if students have learned well before going on to new material"; and "The instructor keeps students informed of their progress."
- 7) Evaluation. The Evaluation dimension measures the extent to which students feel the evaluation instruments (e.g., papers, examinations) fairly assess their ability. Sample items are: "Examinations cover the important aspects of the course"; and "Papers are graded fairly."